Лекция 1 – АиВД

<https://elib.pnzgu.ru/files/eb/imnVd5GIUwhw.pdf>

1. **Введение**

**1.1 Проблема современного общежития**

Мы живем в мире, страдающем от **переизбытка данных**. Веб-сайты отслеживают любое нажатие любого пользователя.

Смартфоны накапливают сведения о вашем местоположении и скорости в ежедневном и ежесекундном режиме. "Оцифрованные" селферы носят шагомеры на стероидах, которые не переставая записывают их сердечные ритмы, особенности движения, схемы питания и сна.

Умные авто собирают сведения о манерах вождения своих владельцев, умные дома — об образе жизни своих обитателей, а умные маркетологи — о наших покупательских привычках.

Сам Интернет представляет собой огромный граф знаний, который, среди всего прочего, содержит обширную гипертекстовую энциклопедию, специализированные базы данных о фильмах, музыке, спортивных результатах, игровых автоматах, мемах2 и коктейлях... и слишком много статистических отчетов (причем некоторые почти соответствуют действительности!) от слишком большого числа государственных исполнительных органов, и все это для того, чтобы вы объяли необъятное. В этих данных кроятся ответы на бесчисленные вопросы, которые никто даже не думает задавать.

**Зачем всё это? – Для обеспечения лучшего существования индивида, в конечном счёте для выживания в условиях конкуренции.**

**1.2 Что такое наука о данных и практика работы с ними**

Наука о данных — это **практическая дисциплина**, которая занимается изучением **методов обобщаемого извлечения знаний из данных**. Она состоит из различных составляющих и основывается на методах и теориях из многих областей знаний, включая обработку сигналов, математику, вероятностные модели, машинное и статистическое обучение, программирование, технологии данных, распознавание образов, теорию обучения, визуальный анализ, моделирование неопределенности, организацию хранилищ данных, а также высокоэффективные вычисления с целью **извлечения смысла из данных** и создания продуктов обработки данных.

**1.3 Направления науки о данных**

**1.3.1** **Data Science и аналитика данных: в чём сходства и различия**

Существует шутка, что аналитик данных — это тот, кто знает **статистику** лучше, чем специалист в области информатики, а информатику — лучше, чем специалист в области статистики.

В значительной мере неважно, как определять понятие науки о данных, потому что всегда можно найти практикующих аналитиков данных, для которых это определение будет всецело и абсолютно неверным.

Можно сказать, что аналитик данных — это тот, кто извлекает **ценные наблюдения из запутанных данных**. В наши дни мир **переполнен** людьми, которые пытаются превратить данные в **ценные наблюдения**.

Data Scientist — это специалист, который создаёт инструменты для решения задач бизнеса. Для этого он использует навыки анализа данных и построения моделей машинного обучения (англ. Machine Learning). Data Scientist работает на стыке трёх областей знания: статистики, машинного обучения и программирования (рисунок 1.1)

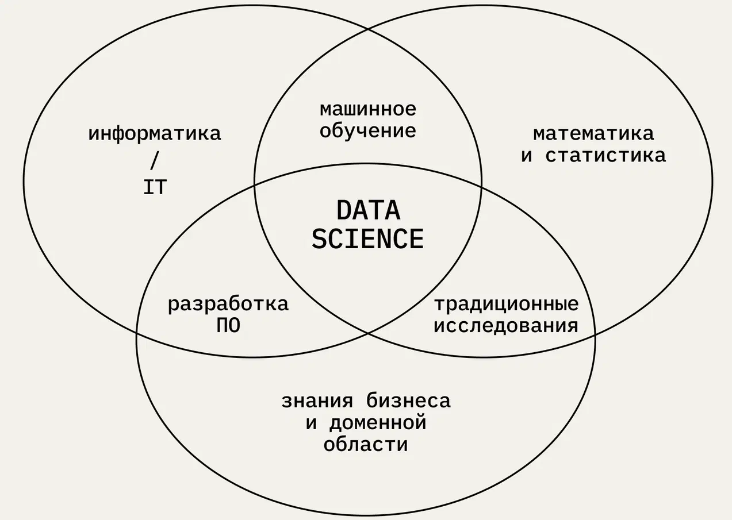


Рисунок 1.1. Области наук, знаний и навыков для профессии «аналитик данных» и «Data Science».

Специалист по Data Science нужен практически в любой индустрии — от промышленности до стриминговых сервисов. Например, в ретейле аналитик данных **анализирует данные о поведении покупателей в магазине и создаёт модель подбора оптимальной цены.**

**В результате средний чек увеличивается, и** **компания получает прибыль**.

Суть работы специалиста по Data Science — **использовать алгоритмы, которые уже придумали до него**, и **разбираться в том, какие из них и когда применять**.

Например, ИИ Netflix, который предлагает сериалы и фильмы на основании того, что люди лайкают и смотрят, — это тоже результат работы дата-сайентиста.

**Суть аналитики в том, чтобы визуализировать данные и делать выводы на их основе**.

Data Science включает в **себя оба эти этапа плюс машинное обучение** (англ. Machine Learning) и глубокое обучение (англ. Deep Learning), с помощью которого специалисты строят и запускают модели для прогнозов.

Обе области пытаются предсказать, как данные повлияют на показатели компании, и предложить лучшие гипотезы, как корректировать поведение пользователей, чтобы достичь результатов. Например, какому сегменту клиентов точно стоит отправлять рассылку, какой отклик планируется получить и окупится ли вложенный в это бюджет. **Аналитика данных и Data Science являются разными этапами одного процесса**.

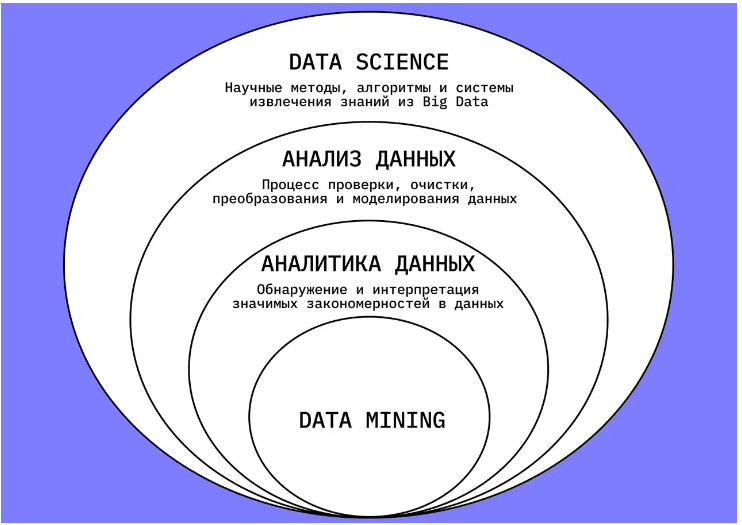


Рисунок 1.2. Взаимосвязь областей науки о данных и технологий обработки данных.

* + 1. **Что такое Data Mining**

Data Mining считается отдельной дисциплиной в области науки о данных.

Data Mining (добыча данных, интеллектуальный анализ данных, глубинный анализ данных или просто майнинг данных)— это процесс, используемый компаниями для **превращения необработанных больших данных в полезную информацию**. Также для этой технологии используется менее популярный термин «обнаружение знаний в данных» или KDD

**KDD -knowledge discovery in databases**.

**Если термином Big Data обозначают все большие данные**— как обработанные, так и нет, то Data Mining представляет собой процесс **глубокого погружения в эти данные для извлечения ключевых знаний**. Это процесс обнаружения в **сырых данных** ранее **неизвестных**, **нетривиальных**, **практически полезных и доступных интерпретации зна**ний, необходимых для принятия решений в различных сферах человеческой деятельности.

**Используя программное обеспечение** для поиска закономерностей в больших пакетах данных, предприятия могут выстраивать маркетинговые стратегии, управлять кредитными рисками, обнаруживать мошенничество, фильтровать спам или даже выявлять настроения пользователей.

**Интеллектуальный анализ данных зависит от эффективного сбора, хранения и компьютерной обработки данных**.

**Первые системы Data Mining** предназначались для обработки данных о продажах в супермаркетах по нескольким параметрам, включая их объем по регионам и тип продукта.

**1.3.3. Задачи Data Mining**

Модели интеллектуального анализа данных применяются для **нескольких типов задач**:

* **прогнозирование**: оценка продаж, предсказание нагрузки сервера или его времени простоя;
* **риск и вероятность**: выбор подходящих заказчиков для целевой рассылки, определение точки баланса для рискованных сценариев, назначение вероятностей по диагнозам или другим результатам;
* **рекомендаци**и: определение продуктов, которые будут продаваться вместе, создание рекомендательных сообщений;
* **поиск последовательностей**: анализ выбора заказчиков во время совершения покупок, прогнозирование их поведения;
* **группирование**: разделение заказчиков или событий на кластеры, анализ и прогнозирование общих черт этих кластеров.

Интеллектуальный анализ данных в **основном используется отраслями, обслуживающими потребителей**, в том числе в сфере розничной торговли, в финансах и маркетинге. Например, у Сбера существует сервис «Сбор Аналитика», который предоставляет данные по отраслям рынка или территориям на основе анализа денежных потоков населения, продаж товаров и услуг и прочих параметров. Его могут использовать как компании, так и госорганы, чтобы оценить потенциал развития региона.

Найти в интернете и рассмотреть применение интеллектуального анализа данных в торговле, в банках и телекоммуникациях, страховании, производстве (система Jet Galatea и другие), социологии и медицине.

**1.4 О современных системах анализа данных**

**1.4.1** **Зачем бизнесу Системы анализа данных**

Программное обеспечение для анализа данных - это решения, обеспечивающие обработку различных бизнес-данных. Такие системы и сервисы позволяют выделять ключевые особенности изучаемого объекта или процесса для пользователя, а также обнаруживать новые возможности продукта, маркетинговые сегменты, отраслевые вертикали и многое другое.

Анализ данных определяется **как процесс автоматизированного/автоматического сбора, очистки, преобразования и моделирования данных для обнаружения полезной информации и принятия управленческих бизнес-решений**.

Всякий раз, когда принимается очередное деловое решение в повседневной жизни, мы думаем, что произошло в прошлый раз, что будет происходить в дальнейшем.

На основании истории и прогноза, ищется и выбирается конкретное решение. Для выполнения анализа могут быть собраны воспоминания о прошлом или планы на будущее, что тоже есть не что иное, как сбор и обработка данных. Все эти операции анализа данных в современный век развития информационных технологий могут выполняться с применением специализированного программного обеспечения - **систем анализа данных**, и с использованием дополнительных источников данных, открытых баз данных, баз знаний и т.п.

**В целом, данный процесс, выполняемый аналитиком, руководителем или предпринимателем для отдельно взятой бизнес-цели, называется анализом данных.**

Анализ данных может быть направлен на достижение различных целей, стоящих перед бизнесом, и иметь различную глубину разработки вопроса:

* **Описательный анализ** – описание процесса КАК ЕСТЬ. Фактическое изложение процесса и проблем, которые необходимо решить. Возможность использования формализованных моделей для передачи знаний между сотрудниками во всю широту компании.
* **Диагностический анализ** – выявление первопричин проблем. Глубокий и детализированный аналитический подход, использующий методы интеллектуального анализа данных для выявления контекста и первопричин бизнес-проблемы.

1. **Прогностический анализ** - прогнозирование на основе исторических данных того, как может развиваться ситуация. Методы, называемые также расширенной аналитикой, используют интеллектуальный анализ данных, машинное обучение и прогнозное моделирование сценариев.

* **Прескриптивный анализ** – формирование видения последующих действий. Используя как исторические данные, так и внешнюю информацию, продукт анализа данных может обеспечить расчет необходимых показателей для выстраивания следующих шагов, которые бизнес должен предпринять для решения заявленных проблем.

**1.4.2 Обзор современных систем анализа данных**

<https://soware.ru/categories/data-analysis-systems/data-source-connectors#productOfCategory>

[**Администрирование**](https://soware.ru/categories/data-analysis-systems/administrations)

Возможность администрирования позволяет осуществлять настройку и управление функциональностью системы, а также управление учётными записями и правами доступа к системе.

[**Анализ больших данных**](https://soware.ru/categories/data-analysis-systems/big-data-analysis)

Функции Анализа больших данных (англ. Big Data Analysis, BDA) реализуют поддержку очень больших наборов данных для исследования предметной области, построения сложных моделей обработки данных и выявления неявных тенденций

[**Визуализация данных**](https://soware.ru/categories/data-analysis-systems/data-visualisation)

Функции Визуализация данных позволяет пользователям выявлять причинно-следственные связи событий, формировать гипотезы или проверять идеи на основании визуального анализа данных

[**Импорт/экспорт данных**](https://soware.ru/categories/data-analysis-systems/importExport)

Возможность импорта и/или экспорта данных в продукте позволяет загрузить данные из наиболее популярных файловых форматов или выгрузить рабочие данные в файл для дальнейшего использования в другом ПО.

[**Индикация трендов и проблем**](https://soware.ru/categories/data-analysis-systems/trend-and-problem-indicators)

Функции Индикации трендов и проблем позволяют пользователям настроить автоматическое определение интересующих событий исходя из набора признаков и факторов

[**Интеллектуальный анализ данных (ИАД)**](https://soware.ru/categories/data-analysis-systems/data-mining)

Функции Интеллектуального анализа данных (ИАД, англ Data Mining, DM) реализуют поиск неочевидных закономерностей, тенденций или извлечения иной информации из больших наборов данных с помощью графических или других инструментов

[**Машинное обучение**](https://soware.ru/categories/data-analysis-systems/machine-learning)

Функции Машинного обучения (англ. Machine Learning, ML) позволяют использовать для решения поставленных задач обучающиеся алгоритмы, проводя исследования на множестве аналогичных заданий, для полной или частичной автоматизации процессов принятия решений, управления рисками и т.д.

[**Многопользовательский доступ**](https://soware.ru/categories/data-analysis-systems/multiUserAccess)

Возможность многопользовательской доступа в программную систему обеспечивает одновременную работу нескольких пользователей на одной базе данных под собственными учётными записями. Пользователи в этом случае могут иметь отличающиеся права доступа к данным и функциям программного обеспечения.

[**Наличие API**](https://soware.ru/categories/data-analysis-systems/existApi)

Часто при использовании современного делового программного обеспечения возникает потребность автоматической передачи данных из одного ПО в другое. Например, может быть полезно автоматически передавать данные из Системы управления взаимоотношениями с клиентами (CRM) в Систему бухгалтерского учёта (БУ). Для обеспечения такого и подобных сопряжений программные системы оснащаются специальными Прикладными программными интерфейсами (англ. API, Application Programming Interface). С помощью таких API любые компетентные программисты смогут связать два программных продукта между собой для автоматического обмена информацией.

[**Отчётность и аналитика**](https://soware.ru/categories/data-analysis-systems/reportAndAnalytic)

Наличие у продукта функций подготовки отчётности и/или аналитики позволяют получать систематизированные и визуализированные данные из системы для последующего анализа и принятия решений на основе данных.

[**Потоковая аналитика**](https://soware.ru/categories/data-analysis-systems/streaming-analytics)

Функции Потоковой аналитики данных позволяют «на лету» применять аналитические алгоритмы над данными в режиме реального времени для отслеживания ключевых показателей бизнес-процессов

[**Прогнозирование и предсказательная аналитика**](https://soware.ru/categories/data-analysis-systems/forecasting-and-predictive-analytics)

Функции Прогнозирования и Предсказательной аналитики позволяют пользователям составлять прогнозы предстоящих затрат, продаж, доходов и иных событий на основании прошлых данных с использованием различных статистических методов прогнозирования

[**Статистический анализ**](https://soware.ru/categories/data-analysis-systems/statistical-analysis)

Функции Статистического анализа дают пользователю инструментарий по математической организации данных, их исследованию, математической интерпретации и представлении данных, а также о выявлении регулярных закономерностей и тенденций

[**Интерактивная аналитическая обработка (OLAP)**](https://soware.ru/categories/data-analysis-systems/olap)

Интерактивная аналитическая обработка (англ. OLAP) позволяет пользователям в реальном времени (онлайн) оперативно получать агрегированную информацию на основе больших массивов данных

[**Коннекторы для источников данных**](https://soware.ru/categories/data-analysis-systems/data-source-connectors)

Коннекторы для источников данных подразумевает либо преднастроенную интеграцию со сторонними источниками данных, либо возможность настройки данного взаимодействия на основе гибкого прикладного программного интерфейса (англ. Application Programming Interface, API)

**1.5 Понятие наборов данных и их источники**

**Сбор высококачественных данных** представляет собой важную часть любого **анализа данных** или **проекта по машинному обучению**. В целях экономии драгоценного времени **команда WebDataRocks** подготовила тщательно отобранный список **бесплатных репозиториев с реальными данными**, готовыми к использованию в проекте.

**1.5.1 Socrata OpenData**

<https://dev.socrata.com/data/>

Одна из крупнейших и наиболее мощных **поисковых систем**, в которой размещены тысячи наборов данных по таким темам, как *финансы, инфраструктура, транспорт, окружающая среда, экономика* и *общественная безопасность*. Все наборы данных категорированы с помощью алгоритмов машинного обучения (рисунок 1.3)

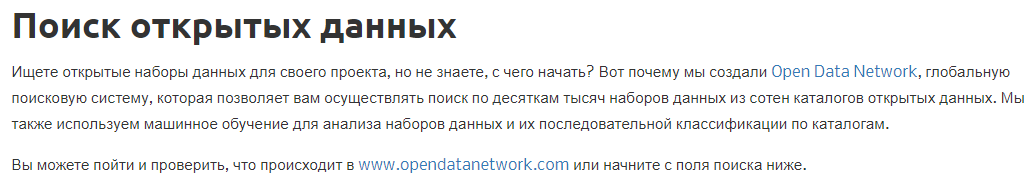


Рисунок 1.3. Перевод страницы сайта <https://dev.socrata.com/data/>

На рисунке 1.4 представлен перевод части страницы сайта <http://www.opendatanetwork.com/>

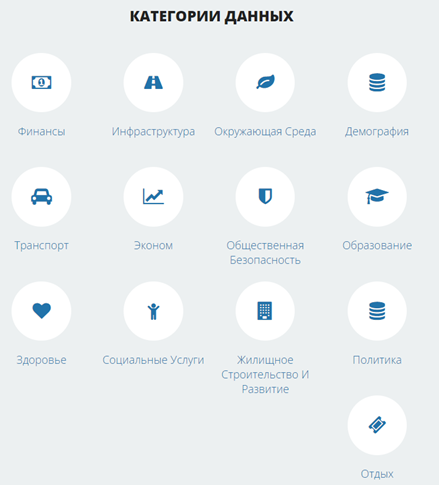


Рисунок 1.4. Иллюстрация возможностей доступа к данным в сети opendatanetwor (переведено на русский)

**Задача** – Ознакомиться со структурой предлагаемых наборов данных (по выбору)

**1.5.2 Kaggle** <https://www.kaggle.com/>

Kaggle (рисунок 1.5) является самой большой платформой и сообществом по науке о данных, которая поражает разнообразием **наборов данных, соревнований** и примеров проектов. Помимо образовательных целей, она предоставляет возможность получить финансовое вознаграждение в соревнованиях, проводимых компаниями, которые стремятся к лучшему пониманию данных. Однако *главная цель этих соревнований — исследование области науки о данных, а не завоевание первых мест*.

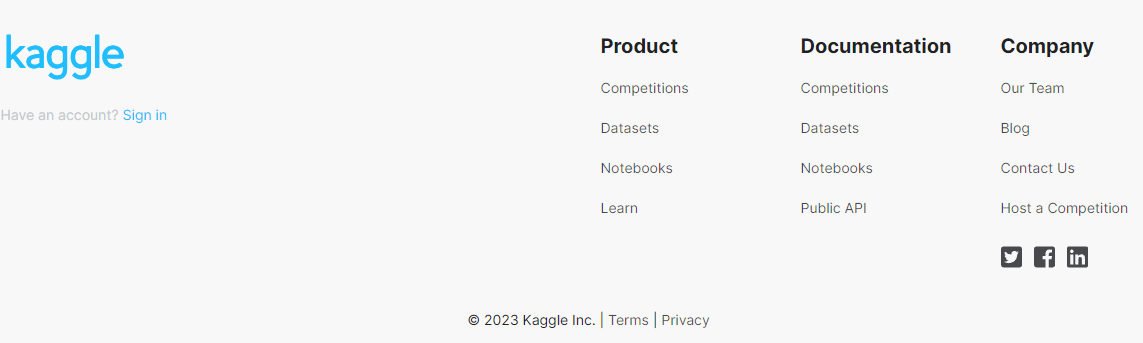


Рисунок 1.5. Фрагмент входной страницы сайта Kaggle

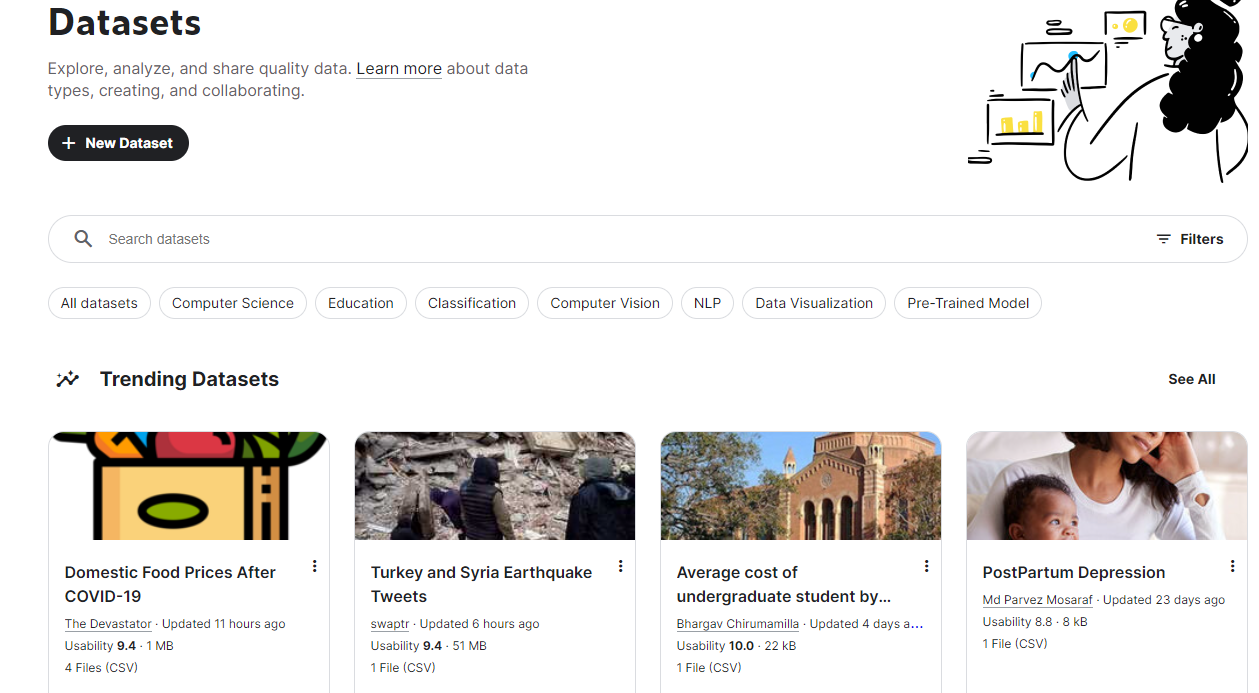


Рис 1.6. фрагмент страницы сайта <https://www.kaggle.com/datasets>

**1.5.3 Следующие интересные наборы датасетов**

* <https://github.com/cvdfoundation/google-landmark>
* [Berlin Airbnb Data](https://www.kaggle.com/brittabettendorf/berlin-airbnb-data)
* [Google-Landmarks Dataset](https://www.kaggle.com/google/google-landmarks-dataset)
* [World Happiness](https://www.kaggle.com/PromptCloudHQ/world-happiness-report-2019)
* [Report 2019](https://www.kaggle.com/PromptCloudHQ/world-happiness-report-2019)
* <http://www.machinelearning.ru/wiki/index.php?title=UCI>
* <https://data.world/resources/knowledge-hub/> !!!!!
* <https://datahub.io/collections>. Имеется множество способов интеграции набора данных в используемый инструмент. Здесь есть фрагменты кода, которые показывают, как использовать данные с **R, Pandas, Python, JavaScript, cURL** и **data-cli**. Кроме того, можно просто загрузить наборы данных **CSV** или **JSON**.
* <https://data.nasdaq.com/> <https://data.nasdaq.com/search?filters=%5B%22Free%22%2C%22Real%20Estate%22%5D>
* <https://data.worldbank.org/>
* <https://data.gov/>
* <https://vc.ru/dev/68180-podborka-datasetov-dlya-mashinnogo-obucheniya>
* <https://otus.ru/nest/post/1377/>

**1.5.4 Открытые данные России**

* <https://data.gov.ru/> (рисунок 1.9),
* <https://ict.moscow/news/datasets/>
* <https://data-in.ru/data-catalog/?order_by=-data_updated_at>
* <https://minzdrav.gov.ru/opendata>
* <https://tproger.ru/translations/the-best-datasets-for-machine-learning-and-data-science/>

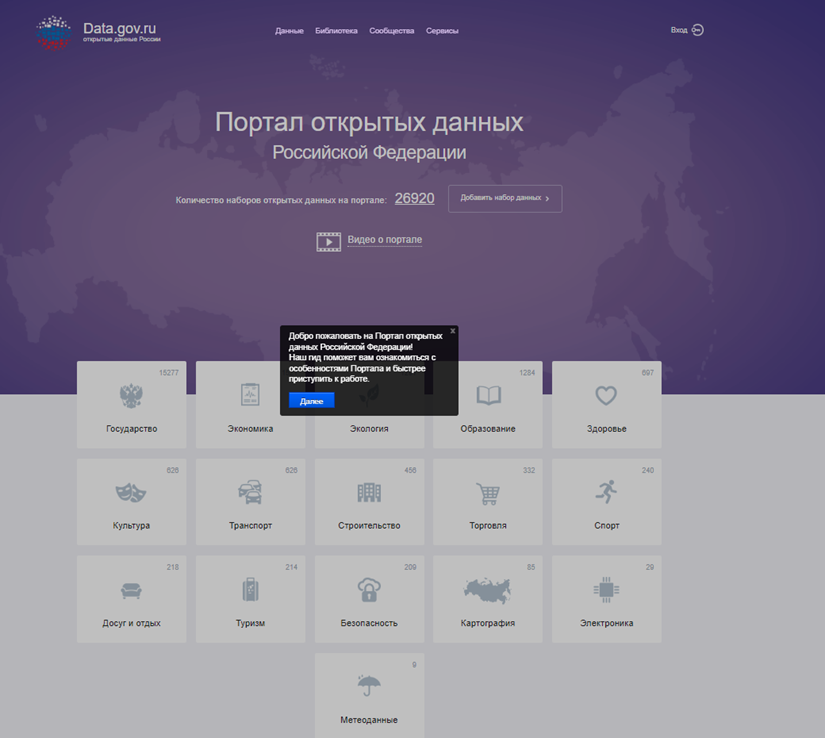


Рисунок 1.7. Фрагмент страницы входа в сайт открытых данных России.

**1.5.4 Что такое dataset, и как его готовить и использовать**

* <https://www.bigdataschool.ru/blog/dataset-data-preparation.html> !!!!!

**1.6 Аналитические системы принятия решений**

В современном мире успех компании на рынке напрямую зависит от того**, как быстро менеджмент компании может распознать изменения динамики рынка и насколько своевременно может отреагировать на них с целью увеличения прибыли**, **исходя из наблюдения за этим рынком**.

**Менеджеры компании должны**:

* отслеживать тенденции рынка,
* идентифицировать конкурентов и угрозы,
* оценивать риски,
* оценивать свои ресурсы и т.д.

Информация является необходимым производственным ресурсом для принятия эффективных управленческих решений. Компании накопили гигантские объемы внутренних и внешних данных. Менеджерам необходимо, чтобы соответствующая информация была соответствующим образом организована для быстрого доступа, анализа и принятия решений.

**Business Intelligence** (сокращённо BI) – это методы и инструменты для поиска, анализа, моделирования и доставки информации, необходимой для принятия решений. Технологии BI обрабатывают большие объемы данных, чтобы найти стратегические возможности для бизнеса.

Рождение BI датируется 1958 годом, когда американский ученый Ханс Петер Лун (1896-1964) опубликовал в IBM System Journal статью «A Business Intelligence System». В ней он **представил бизнес как набор различных видов деятельности в науке, технологиях, коммерции, индустрии и даже в законодательной сфере**, а обеспечивающие его системы – системами, поддерживающими разумную деятельность (intelligence system). Этим словом «intelligence»

**Лун обозначал способность устанавливать взаимосвязь между представлениями отдельных фактов и действиями в интересах решения поставленных задач и достижения намеченных целей**.

В 1989 году аналитик из Gartner Ховард Дреснер дал BI расширительную трактовку, предложив использовать **BI в качестве общего термина для различных технологий, предназначенных для поддержки принятия решений**.

Поддержка BI рассматривается как совокупность различных технологий. **Среди них по-прежнему остается и классический инструмент – электронные таблицы, а также генераторы отчетов**, технологии **OLAP** (**OnLine Analytical Processing** – оперативная аналитическая обработка данных), технологии разработки данных и текстов, а также многое другое. Инструменты BI – программное обеспечение, которое позволяет бизнес пользователям видеть и использовать большое количество сложных данных.

Знания, основанные на данных, (data-based knowledge) получаются из данных с использованием инструментов business intelligence и процесса создания и ведения хранилища данных (**data warehousing**).

Для анализа данных созданы соответствующие программные продукты. В целом сложился рынок инструментальных средств создания и поддержки OLAP-систем, информационных хранилищ (DWH), СППР (DSS), интеллектуального анализа Data mining (DM), который получил обобщенное название – Business intelligence (BI).

Многообразие представленных на рынке решений, от мощных платформ до простых систем аналитики и отчетности, позволяет выбрать решение, доступное любой организации. **Развитие средств визуального представления данных, мобильных и облачных технологий сделали BI-инструменты массовыми всего за последние несколько лет**.

Крупнейшие поставщики предоставляют всевозможные решения для реализации аналитических систем, такие как SAP Business Objects (разработчик – компания SAP AG), Oracle OLAP (разработчик – Oracle Corporation), IBM Smart Analytics System, Statistical Analysis System, Microsoft, Prognoz Platform (разработчик – компания «Прогноз»[[1]](#footnote-1)), АП Loginom (разработчик – компания ООО «Аналитические технологии» Loginom Company) и др.

Компания SAP AG занимается разработкой и внедрением автоматизированных систем управления такими внутренними процессами предприятия, как: бухгалтерский учет, торговля, производство, финансы, управление персоналом, управление складами и т. д. Приложения обычно можно адаптировать под правовой контекст определённой страны. Аналитические приложения SAP работают с разнообразными источниками данных и ИТ-средами, в них предустановлены инструменты управления данными для той или иной отрасли или сектора.

Oracle (Oracle Corporation) – американская корпорация, крупнейший в мире разработчик программного обеспечения для организаций. Oracle Business Intelligence Suite[[2]](#footnote-2) – открытое, основанное на стандартах программное обеспечение, предоставляет единую, интегрированную инфраструктуру для бизнес-анализа, включающую комплексный набор продуктов для обработки запросов и проведения анализа, формирования корпоративных отчетов, доступа к средствам анализа с мобильных устройств, использования информационных панелей и порталов, интеграции с Microsoft Office и Excel, управления процессами бизнес-анализа, рассылки уведомлений в реальном времени, мониторинга бизнес-деятельности и множество других возможностей.

Система **IBM Smart Analytics System** предоставляет гибкий набор функциональных возможностей, включая бизнес-анализ, аналитическую отчетность, оценочные таблицы, инструментальные панели, извлечение информации из данных, сервисы Cubing Services (обеспечивающие многомерную визуализацию данных), текстовый анализ, управление хранилищем данных, а также серверную платформу и среду хранения данных. 9 SAS Institute Inc. Statistical Analysis System – американская частная компания, разработчик технологического программного обеспечения и приложений класса Business Intelligence, Data Quality и Business Analytics.

**Компоненты аналитики SAS**:

• Прогнозная аналитика и интеллектуальный анализ данных (Data Mining) – позволяет строить описательные и прогнозные модели, редактировать их и интегрировать результаты во все бизнес-процессы организации в единой среде.

• Визуализация данных – Повышает эффективность аналитики с помощью динамической интерактивной визуализации данных.

• Прогнозирование и эконометрика – дает возможность анализировать и предсказывать будущие результаты на основе исторических тенденций. • Управление и мониторинг аналитических моделей – значительно упрощает процесс создания аналитических моделей, их управления и передачи на регламентное применение.

• Исследование операций и оптимизация – предлагает методы оптимизации и планирования для поиска наилучшей стратегии решения бизнес-задач.

• Контроль качества – наблюдение процессов и измерение качества процессов во времени.

• Статистика – позволяет использовать статистический анализ данных для принятия решений на основе исторических фактов.

• **Текстовая аналитика** – дает возможность максимально использовать информацию, скрытую в огромных объемах неструктурированных данных.

**Microsoft поставляет BI-инструменты в составе четырёх групп продуктов** – Excel, SharePoint и SQL Server, Power Bi. В каждом из них имеются свои функции аналитики и коллективной работы.

Подведём итог.

**Аналитические платформы** – **это специализированные программные решения, которое содержат все инструменты, необходимые для осуществления процесса извлечения скрытых зависимостей и закономерностей из больших массивов данных,** **предназначенные для формирования отчётности, управления бизнес-процессами, моделирования и прогнозирования показателей, визуализации и оперативного анализа данных, создания бизнес-приложений, в том числе веб приложений и мобильных приложений**.

В аналитических платформах присутствуют гибкие и развитые средства консолидации, включающие интеграционные механизмы с промышленными источниками данных, инструменты очистки и преобразования структурированных и неструктурированных данных, их последующее хранение в едином источнике – **хранилище данных**. Модели, описывающие выявленные зависимости и закономерности, правила, прогнозы также хранятся в специальном репозитории моделей.

**Аналитическая платформа обычно состоит из следующих компонентов:**

• аналитический сервер;

• клиентское приложение;

• подсистема управления метаданными;

• репозитарий моделей;

**1.7 Интеллектуальная аналитическая платформа Loginom**

Целью дисциплины является знакомство студентов с одним из классов интеллектуальных информационных систем – **систем бизнес-аналитики**.

**1.7.1 Общие возможности системы Loginom**

Аналитическая low-code-платформа Loginom предоставляет инструменты для реализации всех аналитических процессов: от интеграции и подготовки данных до моделирования, развертывания и визуализации (рисунок 1.8).

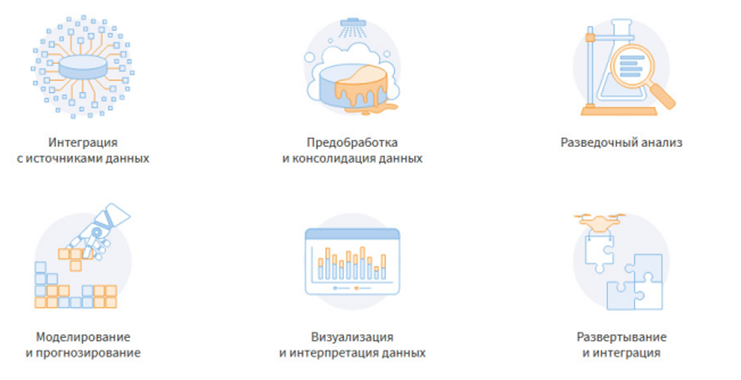


Рисунок 1.8. Процессы аналитики платформы Loginom

Используя платформу Loginom, можно решать следующие бизнес-задачи:

* **управление рисками**:
  + кредитный конвейер, скоринг[[3]](#footnote-3), антифрод[[4]](#footnote-4);
* − **клиентская аналитика**:
  + сегментация клиентов,
  + противодействие оттоку,
  + кросс-продажи;
* **очистка данных**:
  + очистка и удаление дублей, создание золотой записи[[5]](#footnote-5), стандартизация НСИ (нормативно-справочная информация);
* **− маркетинг**:
  + директ-маркетинг, оптимизация цен, оценка эффективности рекламы;
* **логистика**:
  + прогнозирование спроса, оптимизация запасов, расчет страховых запасов;
* − **диагностика**:
  + статистический контроль качества, оценка вероятности поломок, цифровые двойники.

Прикладные бизнес-решения на основе аналитической платформы Loginom представлены на рисунке 1.9



Рисунок 1.9. Прикладные бизнес-решения

Loginom пришла на смену платформе предыдущего поколения – **Deductor**. Платформа Loginom разработана компанией Loginom Company (ранее BaseGroup Labs ООО «Аналитические технологии»).

Ключевые возможности системы:

• Проведение сложных расчетов;

• Консолидация данных;

• Очистка данных;

• Прогнозирование и оптимизация данных;

• Интеграция с любыми хранилищами данных, включая базы данных, отдельные файлы, учетные системы, социальные сети, веб-сервисы.

• Комбинация из структурного и объектно-ориентированного подходов к моделированию.

**Loginom полностью построен на базе веб-технологий**. Веб-интерфейс обладает рядом существенных преимуществ:

• браузеры функционируют на любых ОС (Windows, MacOS, Linux и др.) и множестве устройств;

• минимизируются усилия по администрированию, так как отсутствует необходимость инсталляции и настройки рабочих мест.

Платформа обладает пользовательским интерфейсом, не требующим для работы специальной подготовки. **Loginom имеет поддержку технологий анализа: от простой логики до машинного обучения.**

**В целом в Loginom возможны следующие функции:**

* Визуализация данных.
* Многопользовательский доступ.
* Интерактивная аналитическая обработка (OLAP).
* Администрирование.
* Прогнозирование и предсказательная аналитика.
* Импорт/экспорт данных.
* Отчётность и аналитика.
* Потоковая аналитика.
* Наличие API.
* Машинное обучение.
* Индикация трендов и проблем.
* Статистический анализ.
* Анализ больших данных.
* Коннекторы для источников данных.
* Интеллектуальный анализ данных (ИАД)

**1.7.2** **Запуск и начало работы**

После установки системы Loginom версии Community на ОС Windows можно запуск системы производится щелчком по значку программы в панели запуска (рисунок 1.10).

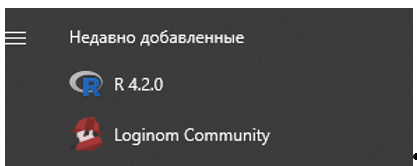


Рисунок 1.10

После запуска приложения (в случае использования настольной редакции) открывается страница «Начало», на которой можно выполнять следующие операции с пакетами (рисунок 1.11):

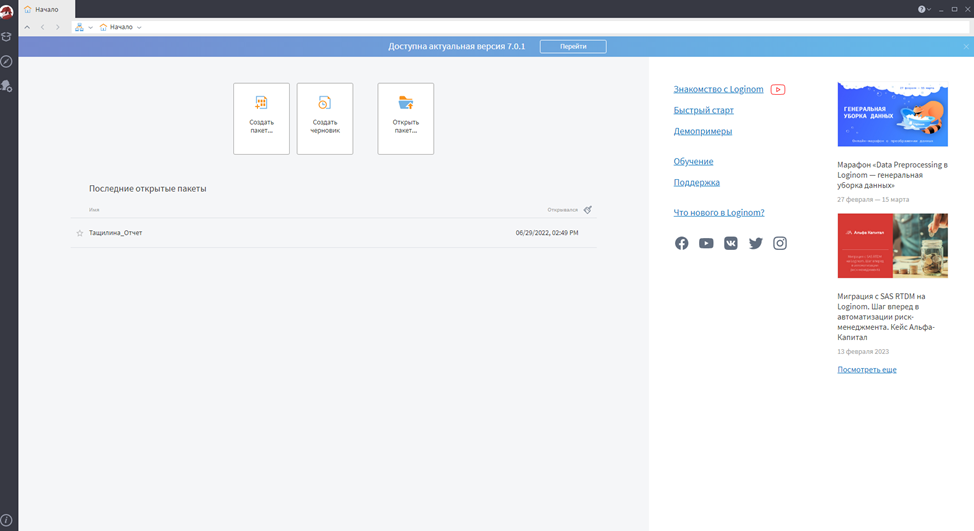


Рисунок 1.11. Интерфейс начала работы с системой Loginom.

* **Создать пакет**— создание нового **пакета**, программа сразу попросить указать путь, куда сохранить пакет.
* **Создать черновик—** создание нового пакета без предварительного сохранения. Сохранить пакет можно будет уже во время работы с ним**.**
* **Открыть пакет—** открытие ранее созданного пакета.

**Важно:** программа не поддерживает автосохранений, при закрытии окна программы (вкладки браузера) все изменения будут утеряны.

После создания пакета, или черновика, или открытия пользователь попадает в **рабочее пространство программы**. Его можно разделить на четыре основные области:

* Крайнее слева главное меню с кнопками: Меню, Пакеты, Навигация, Файлы[[6]](#footnote-6), Процессы[[7]](#footnote-7).
* Верхняя часть отображает вкладки открытых пакетов, содержит адресную строку и элементы для навигации по пакетам и их составляющим.
* Справа от главного меню располагается рабочий стол. Он включает левую панель, где отображаются рабочие компоненты и структура решения (пакеты и их составные части), а также непосредственно область построения сценария и визуализации данных.
* В нижней части окна расположена панель Процессы. По умолчанию она скрыта, но ее можно закрепить (рисунок 1.12).

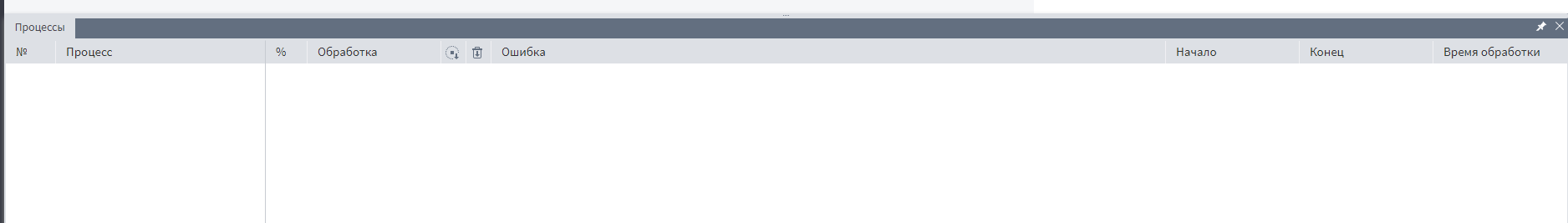


Рисунок 1.12. Раскрытая панель «Процессы»

Стоит отдельно описать **Панель «Процессы»**

Панель «Процессы» предназначена для получения дополнительной информации о процессах обработки данных **узлами** сценариев в рамках текущей сессии Loginom.

В панели иерархически фиксируются процессы и выполняющиеся в них узлы/подузлы, им присваиваются порядковые номера. Каждый новый процесс начинается со строки "Активация узлов", это обусловлено тем, что при старте работы Сценария параллельно могут выполняться сразу несколько узлов (см. рисунок 1.13).

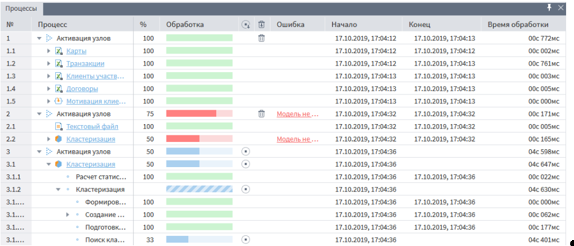


Рисунок 1.13.Панель «Процессы» для работающего сценария

Структура полей панели такова:

* № — номер по порядку.
* Процесс — иерархически указаны процессы (наименования) и их составные части. При нажатии на название процесса в Сценарии будет найден и выделен узел, который отвечает за его выполнение.
* % — процент выполнения процесса.
* Обработка, семантика которой шифруется определённым значком:
  + [https://github.com/loginom/loginom-help/raw/data/interface/information-panel-2.png](https://github.com/loginom/loginom-help/blob/data/interface/information-panel-2.png) — идет процесс обработки, и время его окончания невозможно рассчитать;
  +  — прогресс выполнения текущего процесса по мере обработки данных;
  +  - успешное выполнение (обработка завершена);
  +  - информация о результатах выполнения процесса еще не получена;
  +  - при выполнении процесса произошла ошибка (обработка не завершена);
  +  - выполнение процесса не начиналось, произошла ошибка.
*  - остановить все процессы или
*  - остановить конкретный процесс.
* Ошибка — указывается текст ошибки в случае ее возникновения. При нажатии на данный текст на экран выведется полная формулировка ошибки.
* Начало — дата и время начала процесса.
* Конец — дата и время окончания процесса.

**При выборе процесса на информационной панели можно вызвать контекстное меню, содержащее следующие действия**:

•- Отменить - останавливает выполнение выбранного процесса;

• - Отменить выполнение всех процессов... — останавливает выполнение всех процессов;

• - Удалить из списка — удаляет процесс из списка информационной панели;

• -Удалить все завершённые процессы... — удаляет все завершенные процессы из списка информационной панели;

• **Показать узел** — выделяет выбранный узел на области построения Сценариев;

• **Отображать завершенные процессы** — изменяет отображение процессов на панели. При этом:

При наличии  отображает все процессы, за исключением тех, которые были удалены пользователем;

При отсутствии  показывает только активные процессы и процессы, в которых возникли ошибки. Если в текущий момент времени не выполняется ни один из процессов, а также не было ошибок у процессов завершенных, то в информационной панели данные будут отсутствовать.

**Таким образом с помощью панели «Процессы» можно наблюдать за деталями прохождения расчётов**.

Чтобы при работе со Сценарием информационная панель оставалась открытой, ее нужно закрепить нажатием на кнопку , нажатие на  открепляет панель. Кнопка  панель закрывает.

<https://help.loginom.ru/userguide/workflow/>

<https://examples.loginom.ru/>

<https://www.youtube.com/playlist?list=PLHFj-yQ1Acwm9PJsjJet3VexmvNA7wQwm>

<https://tproger.ru/experts/data-science-practice/>

Социальная сеть - <https://datascientest.com/>

<https://github.com/joelgrus/data-science-from-scratch> Программы по книге Грос

<https://practicum.yandex.ru/blog/kto-takoy-data-scientist/>

<https://www.osp.ru/articles/2022/1004/13056340> !!!!!!!!!!!!!!!!!!!!!!

1. <https://ru.wikipedia.org/wiki/Прогноз_(компания)> [↑](#footnote-ref-1)
2. <https://www.tadviser.ru/index.php/Продукт:Oracle_Business_Intelligence> [↑](#footnote-ref-2)
3. **Кредитный скоринг** (от англ. score «оценка»; в т. ч. гл.) — **система оценки кредитоспособности (кредитных рисков) лица, основанная на численных статистических методах**. [↑](#footnote-ref-3)
4. **Антифрод** (от англ. **anti**-**fraud** «борьба с мошенничеством»), или фрод-мониторинг — система, предназначенная для оценки финансовых и не финансовых событий (карточных транзакций, действий пользователя в ДБО, операций с баллами лояльности и проч.) на предмет подозрительности с точки зрения мошенничества и предлагающая рекомендации по их дальнейшей обработке. [↑](#footnote-ref-4)
5. **Золотая** **запись** — понятие в области управления мастер-данными, которое означает создание наиболее достоверного, проверенного, непротиворечивого и полного представления о каждом объекте данных компании (**клиенте**, товаре, контрагенте и т.д.). **Золотая** **запись** — это единственная, четко определенная версия для каждого объекта данных в информационной системе компании. [↑](#footnote-ref-5)
6. В редакциях Enterprise, Standard и Team (включающих в комплект поставки сервер Loginom) все пользовательские файлы хранятся на сервере Loginom. Таким образом, перед тем как использовать какой-либо файл в сценарии, его необходимо предварительно поместить в файловое хранилище на сервере. Для работы с файловым хранилищем в этих редакциях предусмотрен файловый менеджер.

   **Примечание:** При работе в редакции Personal и Community файловый менеджер отсутствует, так как это версии программ, предназначенные для персональной обработки, но пользователь имеет доступ ко всему файловому пространству локального компьютера.

   При работе с сервером Loginom каждый из пользователей имеет личную папку в файловом хранилище. Физически все папки пользователей располагаются на сервере по следующему пути: "C:\ProgramData\BaseGroup\Loginom 6\UserStorage". Этот путь можно изменить в конфигурационных файлах сервера Loginom. Кроме доступа к личной папке, администратором сервера пользователю могут быть предоставлены права на доступ к другим папкам файлового хранилища. [↑](#footnote-ref-6)
7. Панель **Процессы** может открываться нажатием на кнопку  главного меню. Чтобы при работе со Сценарием информационная панель оставалась открытой, ее нужно закрепить нажатием на кнопку , нажатие на кнопку  открепляет панель. Кнопка   панель закрывает. [↑](#footnote-ref-7)